

Data S1

Note 1. All Protein Observations in MassIVE-KB

The latest versions of the MassIVE-KB are available at

<http://massive.ucsd.edu/ProteoSAFe/static/massive-kb-libraries.jsp>.

All provenance information is accessible for each MassIVE-KB in this link:

Download	Library Name	Description	Precursors	Peptides	Proteins
MGF sptxt	Human HCD Spectral Library	This is the Human HCD spectral library. Containing over 2.1M precursors and drawing on over 30TB of MS/MS proteomics data.	2,122,892	1,088,645	19,610

To view all proteins in MassIVE-KB, click on this link:

Proteins Results
[[All Proteins](#)]

All proteins in SwissProt are listed in this view along with statistics of observation in MassIVE-KB.

Specifically, to view all PE2+ proteins by neXtProt, filter the “neXtProt PE” minimum to 2. To determine the observed proteins in MassIVE-KB according to HUPO criteria, filter the “HUPO Non-Overlapping Peptides” to minimum 2.

To observe all HUPO compliant peptides that confirm the existence of each protein, click on “[View HUPO Peptides](#)”.

Augment Library - Extracting - MSV000080544		Hits 1 - 30 out of 40258		Go to	Go	Export Filtered Results		
Select columns								
Filter	Protein	FDR	score	length	UniProt PE	neXtProt PE	HUPO Non-Overlapping Peptides	HUPO Peptides
1	sp Q86YS7 CZCD5_HUMAN	0.000	1078.343	1000	1	1	66	View HUPO Peptides
2	sp Q9Y6Y8 S23IP_HUMAN	0.000	1910.129	1000	1	1	72	View HUPO Peptides
3	sp Q2YWA4 SKOR2_HUMAN	0.000	536.876	1001	1	1	20	View HUPO Peptides
4	sp Q8N283 ANR35_HUMAN	0.000	362.596	1001	2	1	23	View HUPO Peptides
5	sp O14983 ATZAI_HUMAN	0.000	449.944	1001	1	1	25	View HUPO Peptides
6	sp Q7L7X3 TAOK1_HUMAN	0.000	943.023	1001	1	1	36	View HUPO Peptides
7	sp O14771 TODP1_HUMAN	0.000	655.920	1001	1	1	45	View HUPO Peptides

This highlights all the peptides that are of sufficient length, do not differ by single amino acid variant to another peptide in the database, or are ambiguously identified for the given protein. To trace back the provenance information to all candidate replicate spectra in MassIVE-KB click on “All Candidate Spectra”.

Note 2. All Precursor Observations in MassIVE-KB

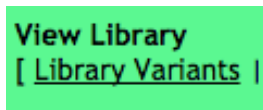
The latest versions of the MassIVE-KB are available at

<http://massive.ucsd.edu/ProteoSAFe/static/massive-kb-libraries.jsp>.

All provenance information is accessible for each MassIVE-KB in this link:

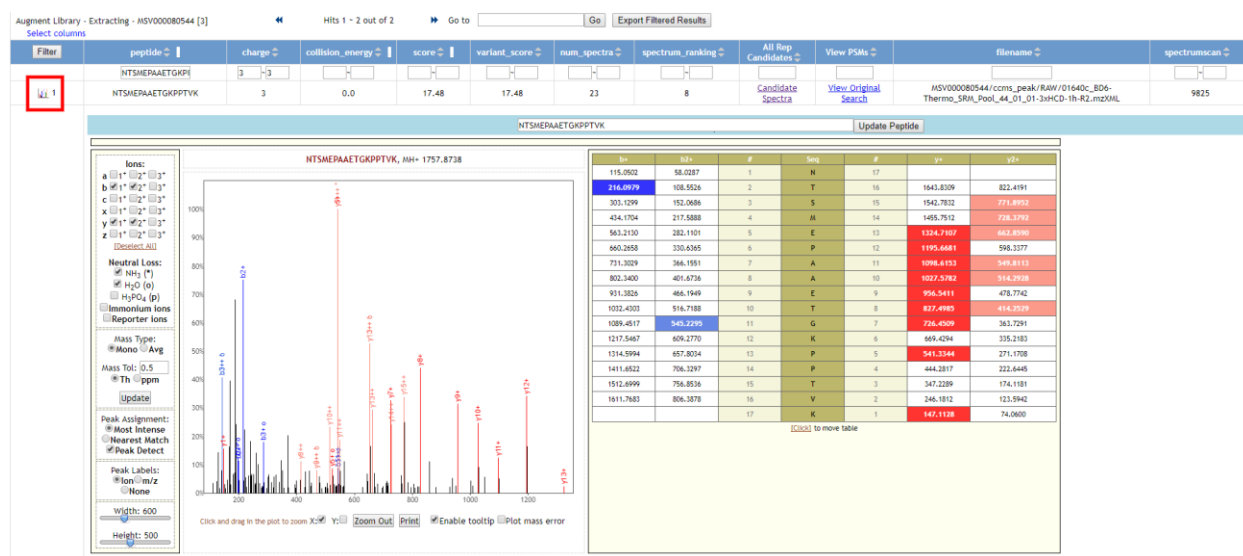
Download	Library Name	Description	Precursors	Peptides	Proteins
MGF sptxt	Human HCD Spectral Library	This is the Human HCD spectral library. Containing over 2.1M precursors and drawing on over 30TB of MS/MS proteomics data.	2,122,892	1,088,645	19,610

To view all precursors in MassIVE-KB, click on this link:



This will list all precursors that were identified and included in MassIVE-KB. To query for a specific peptide, enter the peptide of interest in the “peptide” column. For example: NTSMEPAETGKPPTVK charge 3.

For each entry, we can visualize the spectrum.



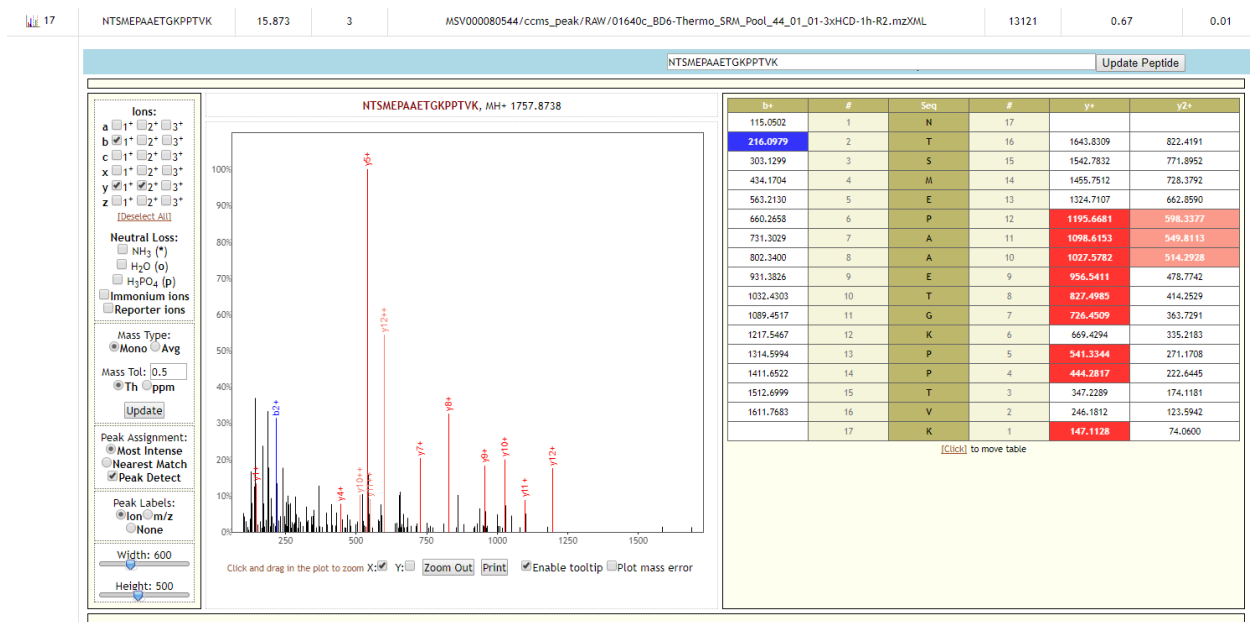
We can find all the spectra in MassIVE where this peptide was identified during MassIVE-KB construction by clicking the “Candidate Spectra” link.

Filter	peptide	charge	collision_energy	score	variant_score	num_spectra	spectrum_ranking	All Rep Candidates	View PSMs	Filename	spectrumscan
1	NTSMEPAETGKPPTVK	3	0.0	17.48	17.48	23	8	Candidate Spectra	View Original Search	MSV000080544/ccms_peak/RAW/D1640c_B04-Thermo_SRA_Pool_44_01_01-3uHCD-1h-R2.mzML	9825
2	NTSMEPAETGKPPTVK	3	0.0	18.08	21.07	20	7	Candidate Spectra	View Original Search	MSV000080544/ccms_peak/RAW/D1866a_GA3-TUM_second_pool_105_02_01-3uHCD-1h-R1.mzML	9653

This gives us the full list of spectra from the data that were in consideration for the representative spectra in MassIVE-KB.

We can see the files where this precursor was found in. Specifically, this precursor was found in both the ProteomeTools Data (Blue Box) as well as in the Bioplex AP-MS data. Precursor was found in the pulldown of the C4orf17 gene, which corresponds to appropriate protein Q53FE4. We can further validate the fragmentation of the spectrum from the AP-MS data:

Against the fragmentation of one of the synthetic spectra:



For each identified peptide spectrum match, we can trace back to the original database search that identified each spectrum by clicking on the “View Original Search”. This will take users back to the status page of the search performed by CCMS.

Note 3. Library Creation Provenance Record

The latest versions of the MassIVE-KB are available at <http://massive.ucsd.edu/ProteoSAFe/static/massive-kb-libraries.jsp>.

All provenance information is accessible for each MassIVE-KB in this link:

Download	Library Name	Description	Precursors	Peptides	Proteins
MGF sptxt	Human HCD Spectral Library	This is the Human HCD spectral library. Containing over 2.1M precursors and drawing on over 30TB of MS/MS proteomics data.	2,122,892	1,088,645	19,610

The construction of MassIVE-KB occurred in several steps.

1. MS/MS data was searched by database tools
2. Identified MS/MS spectra were extracted
3. Extracted MS/MS were incorporated into MassIVE-KB

To view all database searches that were performed on input MassIVE-KB data click on “View All Search Tasks”.

Provenance
[\[All Candidate library spectra](#) [View All Search Tasks](#) [View All Library Augments](#) [View All Library Spectrum Files](#)]

Augment Library - Extracting - MSV000080544 [3]		Hits 1 - 30 out of 7476	Go to	Go	Export Filtered Results
Select columns					
Filter	search_description	View PSMs	Spectra Extraction	Library Augment	
1	FilePartition - 1 of 4 - Bioplex Simple Partition - gene - NKF2.mzXML - Partition 57 of 200	View Original Search	Extract Task	Augment Task	
2	FilePartition - 2 of 4 - Bioplex Simple Partition - gene - RCS01.mzXML - Partition 123 of 200	View Original Search	Extract Task	Augment Task	
3	Kuster Synthetic - Thermo_SRM_Pool_34 - SEARCH - 01649a_B05-Thermo_SRM_Pool_34_01_01-2xIT_2xHCD-1h-R4.mzXML	View Original Search	Extract Task	Augment Task	
Library Augmentation Search - Combined Single Pass Search - MSV000080544 - Run 3 of 10 - PRIDE P90003811 - SILAC-Based Proteomics of Primary Human Kidney Cells treated with Sex Hormones		View Original Search	Extract Task	Augment Task	
5	FilePartition - 3 of 4 - Bioplex Simple Partition - gene - TCTN2.mzXML - Partition 187 of 200	View Original Search	Extract Task	Augment Task	
6	FilePartition - 1 of 4 - Bioplex Simple Partition - gene - THUMP03.mzXML - Partition 69 of 200	View Original Search	Extract Task	Augment Task	
7	FilePartition - 1 of 4 - Bioplex Simple Partition - gene - B3GNT2.mzXML - Partition 80 of 200	View Original Search	Extract Task	Augment Task	
8	FilePartition - 3 of 4 - Bioplex Simple Partition - gene - BCCIP.mzXML - Partition 166 of 200	View Original Search	Extract Task	Augment Task	
9	FilePartition - 3 of 4 - Bioplex Simple Partition - gene - DNAJB8.mzXML - Partition 169 of 200	View Original Search	Extract Task	Augment Task	
10	FilePartition - 3 of 4 - Bioplex Simple Partition - gene - HMG19.mzXML - Partition 154 of 200	View Original Search	Extract Task	Augment Task	
11	Kuster Synthetic - TUM_first_pool_B6 - SEARCH - 01650b_B06-TUM_first_pool_B6_01_01-3xHCD-1h-R2.mzXML	View Original Search	Extract Task	Augment Task	
12	FilePartition - 0 of 4 - Bioplex Simple Partition - gene - FUC42.mzXML - Partition 4 of 200	View Original Search	Extract Task	Augment Task	
13	FilePartition - 1 of 4 - Bioplex Simple Partition - gene - H0RF4L2.mzXML - Partition 75 of 200	View Original Search	Extract Task	Augment Task	
Library Augmentation Search - Combined Single Pass Search - MSV000080596 - Run 27 of 51 - PRIDE PXD004242 - Plasma Proteome Profiling Reveals the Effects of Weight Loss on the Apolipoprotein Family and Systemic Inflammation Status		View Original Search	Extract Task	Augment Task	
15	FilePartition - 0 of 4 - Bioplex Simple Partition - gene - RAB15.mzXML - Partition 40 of 200	View Original Search	Extract Task	Augment Task	
16	FilePartition - 1 of 4 - Bioplex Simple Partition - gene - ASPN.mzXML - Partition 59 of 200	View Original Search	Extract Task	Augment Task	
17	FilePartition - 3 of 4 - Bioplex Simple Partition - gene - PLIN3.mzXML - Partition 150 of 200	View Original Search	Extract Task	Augment Task	
18	FilePartition - 1 of 4 - Bioplex Simple Partition - gene - CORO1C.mzXML - Partition 89 of 200	View Original Search	Extract Task	Augment Task	
19	FilePartition - 0 of 4 - Bioplex Simple Partition - gene - PLAC1.mzXML - Partition 31 of 200	View Original Search	Extract Task	Augment Task	
20	FilePartition - 1 of 4 - Bioplex Simple Partition - gene - FAM719B.mzXML - Partition 78 of 200	View Original Search	Extract Task	Augment Task	
21	FilePartition - 0 of 4 - Bioplex Simple Partition - gene - SEC22B.mzXML - Partition 13 of 200	View Original Search	Extract Task	Augment Task	
22	FilePartition - 2 of 4 - Bioplex Simple Partition - gene - UBL4A.mzXML - Partition 117 of 200	View Original Search	Extract Task	Augment Task	
23	FilePartition - 0 of 4 - Bioplex Simple Partition - gene - CDC4203.mzXML - Partition 14 of 200	View Original Search	Extract Task	Augment Task	
24	FilePartition - 0 of 4 - Bioplex Simple Partition - gene - TRAK2.mzXML - Partition 9 of 200	View Original Search	Extract Task	Augment Task	
25	FilePartition - 3 of 4 - Bioplex Simple Partition - gene - V5K3.mzXML - Partition 183 of 200	View Original Search	Extract Task	Augment Task	
26	FilePartition - 1 of 4 - Bioplex Simple Partition - gene - RNASEH13.mzXML - Partition 79 of 200	View Original Search	Extract Task	Augment Task	
27	FilePartition - 1 of 4 - Bioplex Simple Partition - gene - AKR1C1.mzXML - Partition 65 of 200	View Original Search	Extract Task	Augment Task	
28	FilePartition - 1 of 4 - Bioplex Simple Partition - gene - C4ORV1.mzXML - Partition 84 of 200	View Original Search	Extract Task	Augment Task	
29	FilePartition - 1 of 4 - Bioplex Simple Partition - gene - GSTA3.mzXML - Partition 88 of 200	View Original Search	Extract Task	Augment Task	
30	FilePartition - 2 of 4 - Bioplex Simple Partition - gene - ZPBP2.mzXML - Partition 104 of 200	View Original Search	Extract Task	Augment Task	

Clicking one of these search jobs, takes a user to the status page of the search task. Here users can click on the “Clone” button to visualize all selected parameters in the original analysis. Further users can rerun the identical analysis, facilitating reproducibility.

Job Status	
Workflow	MULTIPASS_MSGF_PLUS_DB_SEARCH
Status	<div> <div> <div>DONE</div> <div>[Clone]</div> </div> <div>[Restart][Delete]</div> </div> <div>[Browse mzTab Result Files]</div> <div> Proteins [View Stats First Pass Proteins Details] </div> <div> PSMs [Rescored PSMs] </div> <div> Precursor Level [First Pass Peptides 1% FDR Second Pass Peptides 1%] </div>

Basic Options

[See here to learn more about MS-GFDB.](#)

Spectrum Files: 2 files and 0 folders are selected

Instrument: Fragmentation Method:

Cysteine Protecting Group: Protease:

Number of Allowed ¹³C: Number of Allowed Non-Enzymatic Termini:

Parent Mass Tolerance: ppm

Allowed Post-Translational Modifications

Maximum Number of PTMs Permitted in a Single Peptide:

	Mass (Da)	Residues:	Type
<input checked="" type="checkbox"/> Oxidation	+15.994915	M	OPTIONAL
<input type="checkbox"/> Lysine Methylation	+14.015650	K	OPTIONAL
<input checked="" type="checkbox"/> Pyroglutamate Formation	-17.026549	Q	OPTIONAL, N-TERMINAL
<input type="checkbox"/> Phosphorylation	+79.966331	STY	OPTIONAL
<input checked="" type="checkbox"/> N-terminal Carbamylation	+43.005814	*	OPTIONAL, N-TERMINAL
<input checked="" type="checkbox"/> N-terminal Acetylation	+42.010565	*	OPTIONAL, N-TERMINAL
<input checked="" type="checkbox"/> Deamidation	+0.984016	NQ	OPTIONAL
<input type="checkbox"/> iTRAQ8plex:13C(6)15N(2)	+304.199040	K	FIXED
<input type="checkbox"/> iTRAQ8plex:13C(6)15N(2)	+304.199040	*	FIXED, N-TERMINAL
<input type="checkbox"/>	+0.984016	NQ	OPTIONAL
<input type="checkbox"/>			<input type="radio"/> FIXED <input checked="" type="radio"/> OPTIONAL <input type="radio"/> FIXED, N-TERMINAL <input type="radio"/> OPTIONAL, N-TERMINAL

Similarly, to examine all MS/MS public data files that were searched, click on “View All Library Spectrum Files”.

Provenance

[\[All Candidate library spectra | View All Search Tasks | View All Library Augments | View All Library Spectrum Files \]](#)

Augment Library - Extracting - MSV000080544 [3]
Select columns

Hits 1 - 30 out of 34684 Go to Go Export Filtered Results

Filter	spectrum_filename	search_description	View Search
1	MSV000078777/ccms_peak/SulfenA_RKO_LCA_A01.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
2	MSV000078777/ccms_peak/SulfenA_RKO_LCA_A02.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
3	MSV000078777/ccms_peak/SulfenA_RKO_LCA_A03.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
4	MSV000078777/ccms_peak/SulfenA_RKO_LCA_B01.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
5	MSV000078777/ccms_peak/SulfenA_RKO_LCA_B02.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
6	MSV000078777/ccms_peak/SulfenA_RKO_LCA_B03.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
7	MSV000078777/ccms_peak/SulfenA_RKO_LCB_A01.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
8	MSV000078777/ccms_peak/SulfenA_RKO_LCB_A02.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
9	MSV000078777/ccms_peak/SulfenA_RKO_LCB_A03.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
10	MSV000078777/ccms_peak/SulfenA_RKO_LCB_B01.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
11	MSV000078777/ccms_peak/SulfenA_RKO_LCB_B02.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
12	MSV000078777/ccms_peak/SulfenA_RKO_LCB_B03.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
13	MSV000078777/ccms_peak/SulfenQ_A431_EGF_LH_A1.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
14	MSV000078777/ccms_peak/SulfenQ_A431_EGF_LH_A2.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
15	MSV000078777/ccms_peak/SulfenQ_A431_EGF_LH_A3.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
16	MSV000078777/ccms_peak/SulfenQ_A431_EGF_LH_B1.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
17	MSV000078777/ccms_peak/SulfenQ_A431_EGF_LH_B2.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
18	MSV000078777/ccms_peak/SulfenQ_A431_EGF_LH_B3.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
19	MSV000078777/ccms_peak/SulfenQ_H202_LH_A1.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
20	MSV000078777/ccms_peak/SulfenQ_H202_LH_A2.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
21	MSV000078777/ccms_peak/SulfenQ_H202_LH_B1.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
22	MSV000078777/ccms_peak/SulfenQ_H202_LH_B2.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search
23	MSV000078777/ccms_peak/SulfenQ_H202_LH_C1.mzXML	Library Augmentation Search - Combined Single Pass Search - MSV000078777 - Run 1 of 1 - Yang: Site-specific mapping and quantification of protein S-sulfenylation in cells	View Original Search

To view all MS/MS extractions run on data and iterative augments to MassIVE-KB click on “View All Library Augments”.

Provenance
[[All Candidate library spectra](#) | [View All Search Tasks](#) | [View All Library Augments](#) | [View All Library Spectrum Files](#)]

This view shows that MassIVE-KB was augmented 10 iterations from 10 extractions of MS/MS data. To examine each augmentation and view a snapshot of MassIVE-KB at a given time point in the past, click on the respective “Library Augment”.

Augment Library - Extracting - MSV000080944 [3]				
Select columns				
Filter	Spectra Extraction	Library Augment	timestamp	
1	Extract_Task	Augment_Task	2017-02-13 11:59:33.0	
2	Extract_Task	Augment_Task	2017-05-19 16:39:08.0	
3	Extract_Task	Augment_Task	2017-01-28 17:54:44.0	
4	Extract_Task	Augment_Task	2017-04-07 10:10:37.0	
5	Extract_Task	Augment_Task	2017-04-05 11:45:02.0	
6	Extract_Task	Augment_Task	2017-02-11 19:02:25.0	
7	Extract_Task	Augment_Task	2017-04-08 11:32:36.0	
8	Extract_Task	Augment_Task	2017-01-02 17:26:50.0	
9	Extract_Task	Augment_Task	2017-04-12 23:44:38.0	
10	Extract_Task	Augment_Task	2017-02-22 21:59:48.0	

Finally, to download full provenance record of all candidate replicate MS/MS spectra for the library, click on the “All Candidate library spectra” link.

Provenance
[All Candidate library spectra](#) | [View All Search Tasks](#) | [View All Library Augments](#) | [View All Library Spectrum Files](#)]

This view contains all top 100 MS/MS replicate spectra per library precursor along with their provenance information, e.g. original database search scores, charge, public dataset file path, scan number, ProteoSAFE search task, explained intensity, similarity to representative, and library creation tasks that brought it into MassIVE-KB. To download in bulk, click the download button on top:

[Back to main page](#)

[Back to status page](#)

[Collapse all](#)

[Download](#)